ABSTRACT

An improved data compression method and apparatus is provided, particularly with regard to the compression of data in tabular form such as database records. The present invention achieves improved compression ratios by utilizing metadata to transform the data in a manner that optimizes known compression techniques. In one embodiment of the invention, a schema is generated which is utilized to reorder and partition the data into low entropy and high entropy portions which are separately compressed by conventional compression methods. The high entropy portion is further reordered and partitioned to take advantage of row and column dependencies in the data. The present invention enables not only much greater compression ratios but increased speed than is achieved by compressing the untransformed data.